

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-356790

(43)Date of publication of application : 26.12.2001

(51)Int.Cl.

G10L 15/02
// G10L101:02
G10L101:065

(21)Application number : 2000-181652

(71)Applicant : CASIO COMPUT CO LTD

(22)Date of filing : 16.06.2000

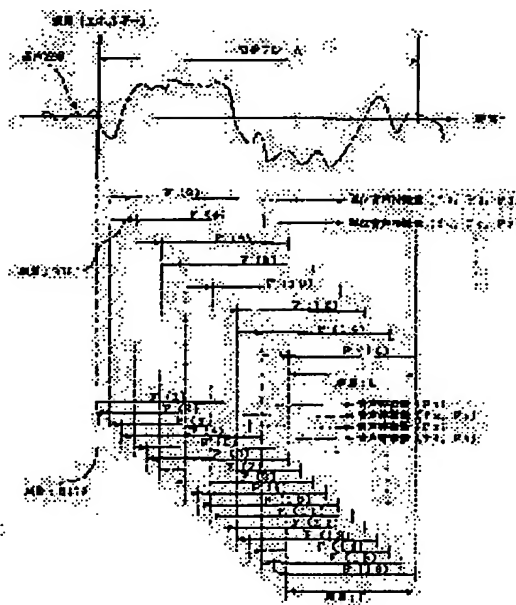
(72)Inventor : KAFUKU SHIGERU

(54) VOICE RECOGNITION DEVICE AND VOICE RECOGNITION METHOD

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a voice recognition device and a voice recognition method in which a voice recognition rate is enhanced without lowering a voice recognition processing speed in a voice recognition device and a voice recognition method making a machine recognize human voice.

SOLUTION: When a featured values extracting part 12 extracts voice featured values from time windows $F(i)$ (i is a natural number) having a fixed length L set for every a cycle $SH2$ from an analysis object voice, the part 12 extracts a frequency axial series-feature parameter concerning the frequency of the voice and a power series-feature parameter concerning the amplitude of the voice in cycles different, respectively.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-356790

(P2001-356790A)

(43) 公開日 平成13年12月26日 (2001. 12. 26)

(51) Int.Cl.⁷

識別記号

F I

テーマコード(参考)

G 1 0 L 15/02

G 1 0 L 101: 02

5 D 0 1 5

// G 1 0 L 101: 02

101: 065

101: 085

7/08

A

5/06

A

審査請求 未請求 請求項の数 4 O L (全 10 頁)

(21) 出願番号 特願2000-181652(P2000-181652)

(22) 出願日 平成12年6月16日 (2000. 6. 16)

(71) 出願人 000001443

カシオ計算機株式会社

東京都渋谷区本町1丁目6番2号

(72) 発明者 加藤 滋

東京都羽村市栄町3丁目2番地1号 カシ

オ計算機株式会社羽村技術センター内

(74) 代理人 100090033

弁理士 荒船 博司 (外1名)

Fターム(参考) 5D015 CC01 CC14 CC18

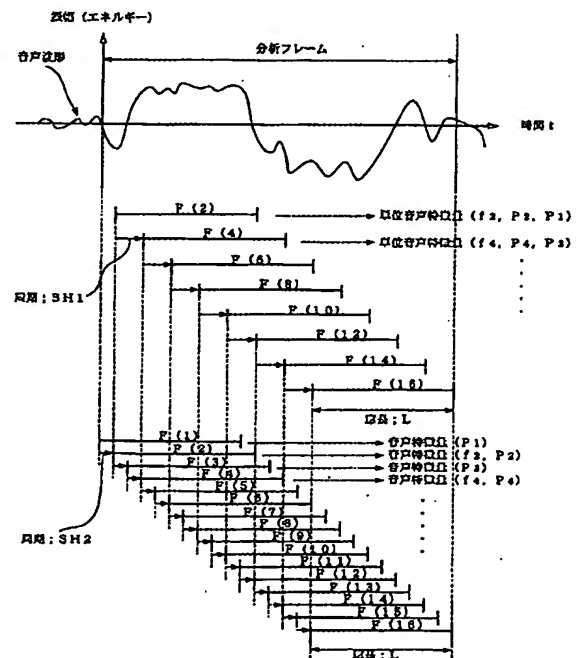
(54) 【発明の名称】 音声認識装置、及び音声認識方法

(57) 【要約】

【課題】 人の音声を機械に認識させる音声認識装置、及び音声認識方法において、音声認識処理速度を低下させることなく、音声認識率を向上させた音声認識装置、及び音声認識方法を提供することである。

【解決手段】 特徴量抽出部12は、分析対象音声から、周期SH2毎に設定された一定長Lの時間窓F

(i) (iは自然数) から音声特徴量を抽出する際には、音声の周波数に関する周波数軸系特徴パラメータと、音声の振幅に関するパワー系特徴パラメータとを、それぞれ異なる周期で抽出する。



【特許請求の範囲】

【請求項 1】分析対象音声に対して所定長の時間窓を所定周期で設定し、この時間窓を処理単位として、前記音声の周波数に関する周波数軸系特徴パラメータと、前記音声の振幅に関するパワー系特徴パラメータとからなる特徴量を抽出する特徴量抽出手段と、

この特徴量抽出手段により抽出された特徴量に基づいて、前記分析対象音声を認識する音声認識手段と、を備えた音声認識装置であって、

前記特徴量抽出手段は、前記周波数軸系特徴パラメータを抽出する周期と、前記パワー系特徴パラメータのみを抽出する周期とを異ならせて、前記特徴量を抽出することを特徴とする音声認識装置。

【請求項 2】前記周波数軸系特徴パラメータを抽出する周期は、前記パワー系特徴パラメータを抽出する周期の自然数倍であることを特徴とする請求項 1 記載の音声認識装置。

【請求項 3】前記音声認識手段は、ある時間窓において抽出された周波数軸系特徴パラメータと、その時間窓において抽出されたパワー系特徴パラメータと、周波数軸系特徴パラメータの抽出周期内であってその時間窓より前に抽出されたパワー系特徴パラメータとを含む特徴量を処理単位として、前記分析対象音声を認識することを特徴とする請求項 1 記載の音声認識装置。

【請求項 4】分析対象音声に対して所定長の時間窓を所定周期で設定し、この時間窓を処理単位として、前記音声の周波数に関する周波数軸系特徴パラメータと、前記音声の振幅に関するパワー系特徴パラメータとからなる特徴量を抽出し、この抽出された特徴量に基づいて、前記分析対象音声を認識する音声認識方法であって、前記周波数軸系特徴パラメータを抽出する周期と、前記パワー系特徴パラメータを抽出する周期とを異ならせて、前記特徴量を抽出することを特徴とする音声認識方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、音声認識装置、及び音声認識方法に係り、詳細には、音声に関する特徴量の抽出方法に特徴のある音声認識装置、及び音声認識方法に関する。

【0002】

【従来の技術】近時、人間の声を機械に認識させるための音声認識の研究が盛んに行われ、部分的には実用化の域に達してきている。音声認識の方法としては、入力音声と、予め与えられた標準パターンとを比較して、最も類似度の大きいものを選び出し、出力する方法が採用されている。

【0003】この標準パターンは、採用する音声認識方法によって異なるものが採用される。例えば、DP (Dynamic Programming; 動的計画法) を用いた DTW (D

ynamic Time Warping; 時間軸非線形マッチング) と呼ばれる音声認識方法においては、典型的な音声特徴量の時系列を標準パターンとして用いている。また、HMM (Hidden Markov Model; 隠れマルコフモデル) を用いた音声認識方法は、統計的手法が駆使されたものであり、音声データに含まれている各単語を音素程度の比較的少ない複数の状態で表し、単語毎に状態の遷移確率と、各状態から入力された特徴量を出力する確率と、をパラメータとして与えられたものを標準パターンとして用いている。現時、この HMM が、中心的な音声認識方法として広く利用されている。

【0004】このような音声認識方式においては、人間が発声した音声信号の中から、音声認識に必要な情報、すなわち音声特徴量を抽出し、標準パターンとの比較を行うが、この音声特徴量の抽出精度の善し悪しが、音声認識全体に関わる性能 (処理速度、及び認識率) の善し悪しを左右する。

【0005】従来、音声入力の音声特徴量を抽出する方法は、図 4 に示すように、分析フレームの範囲内でハミング窓のような一定値 L の窓長を有する時間窓を時系列に一定時間ずつシフト (周期 SH1) させて設定し、設定された各時間窓 F (1) ~ F (8) によって順次切り出された分析フレーム内の音声波形からの音声サンプル (N 標本数点) をそれぞれ取り出し、時間窓の音声サンプルの音声特徴量を抽出する方法を用いる。

【0006】このような音声特徴量は、時間窓毎に順次切り出された所定数の音声サンプルを周波数軸上に変換することによって得られる音声特徴量 (以後、これを周波数軸系特徴パラメータと呼ぶ) と、線形 PCM (Pulse Code Modulation; パルス符号化) により量子化された音声振幅の 2 乗またはその対数をとることによって得られる音声特徴量 (以後、これをパワー系特徴パラメータと呼ぶ) とを組み合わせることによって得られるものである。

【0007】例えば、周波数軸系特徴パラメータ 12 成分 (12 次元) とパワー系特徴パラメータ 1 成分 (1 次元)、及び直前の時間窓の各成分との差分を取ったもの、すなわち Δ 周波数軸系特徴パラメータ 12 成分 (12 次元) と Δ パワー系特徴パラメータ 1 成分 (1 次元) の、合計 26 成分を 26 次元ベクトル量として特徴量を構成する等が考えられる。

【0008】

【発明が解決しようとする課題】しかしながら、このような従来の音声認識方式においては、次のような問題点があった。入力音声の音声特徴量は、上述のように分析フレームの範囲内で時間窓を等間隔にずらして一部重なり合うよう設定して、各時間窓で切り出された音声サンプルを基に抽出される。この時間窓の重なり合う時間を短く設定した場合は、時間窓の数が減少するので入力音声から切り出す音声サンプル数が少なくなり、音声特徴

量の抽出処理回数、及び、その後段に続く音声認識処理回数が共に低く抑えられ、処理速度を上げることが可能となる一方で、サンプリングが粗くなるので統計性が悪くなり、音声認識率は低下してしまう。他方、時間窓の重なり合う時間を長く設定した場合、時間窓の数が増大するのでサンプリングが細くなり入力音声から切り出す音声サンプル数が多くなるので、統計性が向上し、精度の良い音声認識が可能となるが、その反面、音声特徴量の抽出処理回数、及び、その後段に続く音声認識処理回数が共に増大し、処理速度の低下を招いてしまう。

【0009】すなわち、処理速度を上げると認識率が下がり、認識率を高めようとすると処理速度が低下してしまい、処理速度と認識率を同時に向上させることが困難であった。

【0010】本発明の課題は、音声認識処理速度を低下させることなく、音声認識率を向上させた音声認識装置、及び音声認識方法を提供することである。

【0011】

【課題を解決するための手段】請求項1記載の発明は、分析対象音声（例えば図2に示す音声データ $d(n)$ ）に対して所定長（例えば図2に示す窓長 L ）の時間窓（例えば図2に示す時間窓 $F(1) \sim F(16)$ ）を所定周期（例えば図2に示す周期 $SH2$ ）で切り出し、この時間窓を処理単位として、前記音声の周波数に関する周波数軸系特徴パラメータと、前記音声の振幅に関するパワー系特徴パラメータとからなる特徴量を抽出する特徴量抽出手段（例えば図1に示す音声特徴量抽出部12）と、この特徴量抽出手段により抽出された特徴量に基づいて、前記分析対象音声を認識する音声認識手段（例えば図1に示す比較部13）と、を備えた音声認識装置であって、前記特徴量抽出手段は、前記周波数軸系特徴パラメータを抽出する周期（例えば図2に示す $SH1$ ）と、前記パワー系特徴パラメータのみを抽出する周期（例えば図2に示す $SH2$ ）とを異ならせて、前記特徴量を抽出することを特徴とする。

【0012】この請求項1記載の発明によれば、周波数軸系特徴パラメータを抽出する周期と、パワー系特徴パラメータを抽出する周期とを異ならせることによって、最も効率良く音声認識ができるように認識処理速度と認識率を最適化することが可能となり、処理速度を低下させることなく音声認識率を向上させることが可能となる。

【0013】また、請求項2記載の発明のように、請求項1記載の音声認識装置において、周波数軸系特徴パラメータを抽出する周期が、パワー系特徴パラメータを抽出する周期の自然数倍である（例えば、図2に示す $SH1$ と $SH2$ ）、としても良い。

【0014】この請求項2記載の発明によれば、比較的小さい計算量で得られ、かつ音声認識率の向上に有効なパワー系特徴パラメータを抽出する頻度が、周波数軸系

特徴パラメータを抽出する頻度よりも多く（自然数倍）なるので、処理速度を低下させることなく、音声認識率を向上させることが可能となる。

【0015】更に、請求項3記載の発明のように、請求項1記載の音声認識装置において、ある時間窓において抽出された周波数軸系特徴パラメータと、その時間窓において抽出されたパワー系特徴パラメータと、周波数軸系特徴パラメータの抽出周期内であって、その時間窓より前に抽出されたパワー系特徴パラメータとを含む特徴量を処理単位として、前記分析対象音声を認識する（例えば、図2に示す単位音声特徴量）、としても良い。

【0016】この請求項3記載の発明によれば、同一の時間窓において抽出されたパワー系特徴パラメータと周波数軸系特徴パラメータとからなる従来の音声特徴量に対して、更に、それ以前の時間窓において抽出されたパワー系特徴パラメータを付加したものを新たな音声特徴量として用いるので、処理速度を低下させることなく、従来に比べてより精度良く音声認識を行うことが可能となる。

10 【0017】

【発明の実施の形態】以下、図1～図3を参照して本発明の実施の形態における音声認識装置1について詳細に説明する。まず構成を説明する。

【0018】図1は、本発明の一実施の形態におけるHMMモデルを用いた音声認識装置1の内部構成を示すブロック図である。この図1に示すように、音声認識装置1は、時間窓位置設定部11、音声特徴量抽出部12、比較部13、及び記憶装置（図示せず）内に予め格納されている標準パターン（HMMモデル141～HMMモデル14n）とから構成されている。

30

【0019】図2は、時間窓位置設定部11における、入力音声波形に対する時間窓の設定の様子を示す図であり、一例として、分析対象とする音声区間（分析フレーム）から16個の時間窓を切り出す例を示している。

【0020】図2の上段は、入力された音声波形の分析対象とする一定長の音声区間（以下、分析フレームと呼ぶ）を示しており、横軸は時間軸、縦軸は音声波形の振幅（エネルギー）を表している。図2の中段は、分析フレームの範囲内においてシフト長 $SH1$ ずつ等間隔でシフトしてなる8つの時間窓 $F(i)$ （ i は2～16までの偶数）を示している。これらの時間窓は一定長の窓長 L を有し、隣り合う時間窓が一部重複するようになっている。図2の下段は、分析フレームの範囲内で等間隔に設定される16個の時間窓 $F(i)$ （ i は1～16までの自然数）を示している。各時間窓は一定長の窓長 L を有し、隣り合う時間窓が一部重複するようにして、周期 $SH1$ の2倍の周期 $SH2$ で設定されている。

40

【0021】図1に示す時間窓位置設定部11には、図2に示す時間窓の窓長 L 、及び時間窓設定周期 $SH2$ が設定されている。時間窓位置設定部11は、分析フレー

50

ムの範囲内において設定されている窓長 L 、及び設定周期 $SH2$ に従って、 i 番目の時間窓 $F(i)$ を設定し、音声特徴量の抽出を開始させるための開始制御信号を音声特徴量抽出部12に出力する。また、窓長 L の後、時間窓 $F(i)$ における音声特徴量の抽出を終了させるための終了制御信号を音声特徴量抽出部12に出力する。この一連の動作を、周期 $SH2$ 毎に分析フレームが終了するまで繰り返す。

【0022】音声特徴量抽出部12は、前記時間窓位置設定部11から入力される時間窓 $F(i)$ の開始制御信号及び終了制御信号に基づいて、入力音声を分析フレーム内の i 番目の時間窓 $F(i)$ で切り出し、この時間窓 $F(i)$ 内の音声データ $d(n)$ からパワー系特徴パラメータ P_i （1次元ベクトル量）や周波数軸系特徴パラメータ f_i （例えば、 D 次元ベクトル量）を抽出する。

【0023】パワー系特徴パラメータ P_i とは音声の振幅に関する特徴量であり、例えば、音声データ $d(n)$ のエネルギー二乗和やその対数を計算するといった比較的少ない計算量で求められる1次元ベクトル量である。周波数軸系特徴パラメータ f_i とは、例えば、ケプストラム、メルケプストラムと呼ばれる音声の周波数に関する特徴量であり、音声データ $d(n)$ に対してのFT（Fourier Transform；フーリエ変換）、対数変換、メル軸変換等の複数（例えば、 D 個）の計算結果から構成される D 次元ベクトル量である。

【0024】音声特徴量抽出部12は時間窓 $F(i)$ が偶数番目の窓の場合、すなわち、 i が偶数の場合は、パワー系特徴パラメータ P_i とともに周波数軸系特徴パラメータ f_i も抽出する。また、時間窓 $F(i)$ が奇数番目の窓の場合、すなわち、 i が奇数の場合は、パワー系特徴パラメータ P_i のみ抽出する。つまり、図2に示すように、周期 $SH2$ 毎に分析フレームの範囲内から時間窓で入力音声の切り出されるが、パワー系特徴パラメータ P_i は全ての時間窓で抽出され、周波数軸系特徴パラメータ f_i は偶数番目の時間窓だけから抽出される。従って、パワー系特徴パラメータ P_i は、周波数軸系特徴パラメータ f_i の $1/2$ 倍の周期（2倍の頻度）で抽出される。

【0025】また、音声特徴量抽出部12は、 i 番目の時間窓 $F(i)$ （ i は偶数）から抽出される周波数軸系特徴パラメータ f_i （ D 次元ベクトル量）とパワー系特徴パラメータ P_i （1次元ベクトル量）に、直前（奇数）の時間窓 $F(i-1)$ から抽出されるパワー系特徴パラメータ P_{i-1} （1次元ベクトル量）を付加し、これらの特徴パラメータ f_i 、 P_i 、 P_{i-1} を合わせて $D+2$ 次元ベクトル量（ f_i 、 P_i 、 P_{i-1} ）を構成し、このベクトル量を単位音声特徴量として比較部13に出力する。具体的には図2の中段に示すように周期 $SH1$ で単位音声特徴量（ f_i 、 P_i 、 P_{i-1} ）（ i は偶数）を比較部13に出力する。従って、特徴量は周期 $SH1$ で抽出

されたものとみなされて比較部13に入力されるが、一周期 $SH1$ 内にある時間窓（ここでは $i-1$ 番目の時間窓 $F(i-1)$ ）から抽出されたパワー系特徴パラメータ P_{i-1} を付加しているため、音声特徴量は従来と比較して精度よく（細かく）抽出されている。

【0026】更に、隣り合う周波数軸系特徴パラメータの差分 Δf_i （ f_i と同じ次元を持つベクトル量で、例えば D 次元）や隣り合うパワー系特徴パラメータの差分 ΔP_i （1次元ベクトル量）を演算し、これらの差分 Δf_i 、 ΔP_i を単位音声特徴量（ f_i 、 P_i 、 P_{i-1} ）に付加するようにしてもよい。

【0027】比較部13は、前記分析フレーム内で抽出された各種特徴パラメータからなる前記単位音声特徴量と、記憶装置（図示せず）内に予め格納してある標準パターンとを比較照合して、入力音声の認識を行い、認識結果を出力する。本実施例では、HMMに基づく統計的手法により音声認識を行うこととする。

【0028】ここで、上記HMMを利用した統計的手法に基づく音声認識の手法を説明する。音声特徴量抽出部12にて抽出された入力音声の音声特徴量を用いて、予め与えられているHMMモデル141～14nと呼ばれる標準パターンに基づいて音声認識を行う。HMMモデル141～14nとは、音声データに含まれている各単語を音素程度の比較的少ない複数の状態で表し、単語毎に状態の遷移確率と、各状態から入力された特徴量を出力する確率と、をパラメータとして与えたものである。比較部13では、HMMモデル141～14nの中で、どのHMMモデルが、与えられた音声特徴量を最も高い確率で出力するか、を尤度（確率）計算し、その確率を最大とするHMMモデルに対応する単語を音声認識結果として出力する。

【0029】次に動作を説明する。図3は、音声認識装置1の音声認識処理を説明するフローチャートである。

【0030】音声認識装置1に対して音声が入力されると、入力された音声は、入力段に設けられたA/D変換器（図示せず）により、所定のサンプリング間隔でサンプリングして標本化を行う線形PCM方法に基づいてPCM符号化された音声データ $d(n)$ に変換され、時間窓位置設定部11、及び音声特徴量抽出部12に出力される。

【0031】音声データ $d(n)$ が時間窓位置設定部11に入力されると、ハミング窓のような時間窓 $F(i)$ を特定する開始制御信号・終了制御信号が、時間窓位置設定部11から音声特徴量抽出部12へ出力される（ステップS30）。次いで、音声特徴量抽出部12では、ステップS30で出力された前記制御信号に基づいて、時間窓 $F(i)$ における音声データ $d(n)$ が切り出される（ステップS31）。

【0032】音声特徴量抽出部12は、ステップS31で切り出された時間窓 $F(i)$ における音声データ d

(n) の 2 乗和またはその対数をとることによってパワー系特徴パラメータ P_i を計算する (ステップ S 3 2)。

【0033】更に、音声特徴量抽出部 1 2 は、ステップ S 3 2 において計算されたパワー系特徴パラメータ P_i を記憶する (ステップ S 3 3)。特に i が偶数である場合、一段前の時間窓 $F(i-1)$ において得られたパワー系特徴パラメータ P_{i-1} も加えて、パワー系特徴パラメータ P_i 及び P_{i-1} を一組としてまとめて記憶する。

【0034】また、音声特徴量抽出部 1 2 では、ステップ S 3 2 で得られたパワー系特徴パラメータ P_i と、予め設定されたエネルギーレベル (振幅) とを比較することによって、この時間窓 $F(i)$ で切り出された音声区間が有音であるか、無音であるか、または音声区間が終了したか、を判断する (ステップ S 3 4)。

【0035】ステップ S 3 4 において、パワー系特徴パラメータ P_i が前記エネルギーレベルを越えている場合、当該時間窓 $F(i)$ が有音区間と判断して (ステップ S 3 4 ; 有音)、ステップ S 3 5 に進む。

【0036】ステップ S 3 5 において、音声特徴量抽出部 1 2 は、時間窓 $F(i)$ の設定順序 i が奇数であると判断した場合 (ステップ S 3 5 ; No)、そのままステップ S 3 8 に移行し、時間窓 $F(i)$ から時間窓 $F(i+1)$ ヘシフトして、ステップ S 3 1 に戻る。

【0037】一方、前記ステップ S 3 5 において、音声特徴量抽出部 1 2 は、時間窓 $F(i)$ の設定順序 i が偶数であると判断した場合 (ステップ S 3 5 ; Yes)、時間窓 $F(i)$ における音声データ $d(n)$ を、FT 等により周波数軸上に変換することによって得られる周波数軸系特徴パラメータ f_i を抽出し (ステップ S 3 6)、ステップ S 3 3 で記憶されたパワー系特徴パラメータ P_i と共に、時間窓 $F(i)$ の音声特徴量として記憶する (ステップ S 3 7)。次いで、ステップ S 3 8 に移行し、時間窓 $F(i)$ から時間窓 $F(i+1)$ ヘシフトして、ステップ S 3 1 に戻る。

【0038】ステップ S 3 4 の判断において、パワー系特徴パラメータが前記エネルギーレベルに満たない場合は、当該時間窓 $F(i)$ の一段前の時間窓 $F(i-1)$ で取り出された音声データ $d(n)$ の状態に応じて、以後の処理が異なる。すなわち、ステップ S 3 4 において、直前の時間窓 $F(i-1)$ が有音区間のとき、この時間窓 $F(i)$ で音声区間 (分析フレーム) が終了したものと判断して (ステップ S 3 4 ; 音声区間終了)、ステップ S 3 9 へ移行する。一方、直前の時間窓 $F(i-1)$ が無音区間のとき、この時間窓 $F(i)$ では無音区間が継続しているもの、つまり音声はまだ発声されていない状態であると判断して (ステップ S 3 4 ; 無音)、ステップ S 3 8 へ移行し、時間窓 $F(i)$ から時間窓 $F(i+1)$ ヘシフトして、ステップ S 3 1 に戻る。

【0039】ここで i が偶数の場合、ステップ S 3 7 で

記憶される時間窓 $F(i)$ で抽出された音声特徴量には、この時間窓 $F(i)$ で抽出されたパワー系特徴パラメータ P_i と周波数軸系特徴パラメータ f_i と共に、一段前の時間窓 $F(i-1)$ で抽出されたパワー系特徴パラメータ P_{i-1} も付加される。すなわち、時間窓 $F(i)$ (i は偶数) で記憶される音声特徴量とは、この時間窓 $F(i)$ で抽出された周波数軸系特徴パラメータ f_i (例えば D 成分) 及びパワー系特徴パラメータ P_i と、直前の時間窓 $F(i-1)$ で抽出されたパワー系特徴パラメータ P_{i-1} とを、1 つのベクトル量としてまとめたものであり、周波数軸系特徴パラメータ (D 次元) + パワー系特徴パラメータ (2 次元) からなる $D+2$ 次元ベクトル量 (f_i, P_i, P_{i-1}) として記憶される。

【0040】また、時間窓 $F(i)$ (i は偶数) において $D+2$ 次元ベクトル量として記憶される音声特徴量は、二段前の時間窓 $F(i-2)$ (i は偶数) の音声特徴量 ($D+2$ 次元ベクトル量) との差分 ($\Delta f_i, \Delta P_i, \Delta P_{i-1}$) も特徴量として加え、結局、周波数軸系特徴パラメータ (D 次元) + パワー系特徴パラメータ (2 次元) + Δ 周波数軸系特徴パラメータ (D 次元) + Δ パワー系特徴パラメータ (2 次元) からなる $2D+4$ 次元ベクトル量 ($f_i, P_i, P_{i-1}, \Delta f_i, \Delta P_i, \Delta P_{i-1}$) とすることも可能である。

【0041】また、ステップ S 3 4 において、直前の時間窓 $F(i-1)$ が有音区間のとき、当該時間窓 $F(i)$ で音声区間が終了したと判断すると、ステップ S 3 7 において時間窓 $F(i)$ (i は偶数) 毎に記憶された音声特徴量を時系列に配置する (ステップ S 3 9)。

【0042】次いで、ステップ S 3 9 において時系列に配列された音声特徴量は、比較部 1 3 に出力され、予め記憶されている標準パターン (HMM モデル 1 ~ HMM モデル n) と比較照合される (ステップ S 4 0)。ステップ S 4 0 で得られた結果は音声認識装置 1 の図示しない出力段に出力され (ステップ S 4 1)、一連の音声認識処理を終了する。

【0043】以上説明したように、本実施の形態における音声認識装置 1 によれば、音声の分析対象となる分析フレームから時間窓を設定して音声特徴量を抽出する場合に、パワー系特徴パラメータを抽出する時間窓 F

(1) ~ $F(16)$ の設定周期を SH2 とし、周波数軸系特徴パラメータを抽出する時間窓 $F(i)$ (i は偶数) の設定周期を SH の 2 倍の周期である SH1 としている。

【0044】このため、パワー系特徴パラメータを周波数軸系特徴パラメータより多く抽出しているため、音声認識率が向上する。また、パワー系特徴パラメータは比較的少ない計算量で求められるので、音声認識装置 1 の処理負荷を比較的小さく抑えつつ、認識率を向上させることができる。

【0045】また偶数番目 (i は偶数) の時間窓のパワ

一系特徴パラメータ P_i と周波数軸系特徴パラメータ f_i と、その直前（奇数番目）の時間窓のパワー系特徴パラメータ P_{i-1} とをまとめて単位音声特徴量とし、周波数軸系パラメータ f_i の抽出周期 $SH1$ で取り出すようにしているため、比較部 13 における音声認識処理の処理速度は従来と同程度に抑えることができるので、音声認識の処理速度を低下させることなく、音声認識率を向上させることができる。

【0046】なお、本発明は、上記実施の形態の内容に限定されるものではなく、本発明の趣旨を逸脱しない範囲で適宜変更可能であり、例えば、周波数軸系特徴パラメータ f_i を抽出する時間窓の設定周期 $SH1$ は、周期 $SH2$ の 2 倍としているが、周期 $SH1$ を、周期 $SH2$ の N （自然数）倍として自由に設定できるようにしても良い。この際、図 3 におけるステップ S35 では、 i が N の自然数倍であるか否か、を判断する。その際、抽出される音声特徴量は、 $D+N$ 次元ベクトル量（ f_i , P_i , \dots , P_{i-N+1} ）として、ステップ S37 において記憶される。周波数軸系特徴パラメータ f_i に基づく音声認識処理は、より高い精度の認識処理が可能である反面、パワー系特徴パラメータ P_i に基づく音声認識処理よりも多くの計算を必要とするので、適当な $SH1/SH2$ 比を選ぶことにより、音声認識上最も効率の良い認識処理速度と認識率を設定することが可能となる。

【0047】また、上記実施の形態においては、HMM を用いた音声認識としたが、これに限定されるものではなく、例えば、DTW と呼ばれている音声認識方法を用いても良い。

【0048】更に、上記実施の形態においては、分析フレーム内における時間窓の設定周期を $SH2$ に固定してあるが、音声区間外においては、 $SH2$ よりも長い周期に設定することも可能である。この場合、音声区間外における認識処理が簡略化されるので、認識処理の効率化が期待できる。

【0049】また、所定長の分析フレームに対して 16 個の時間窓 $F(i)$ を設定し、比較部 13 に出力する単位音声特徴量は 8 個としたが、時間窓等の個数はこれに限定されるものではなく、例えば、所定長の分析フレームに対して時間窓 $F(i)$ を 14 個設け、従って比較部に出力する単位音声特徴量は 7 個とするなど、適宜変更

可能である。

【0050】

【発明の効果】請求項 1、及び 4 記載の発明によれば、入力音声の特徴量の抽出する際、周波数軸系特徴パラメータを抽出する周期と、パワー系特徴パラメータを抽出する周期とを異ならせることによって、最も効率良く音声認識ができるように認識処理速度と認識率を最適化することが可能となり、処理速度を低下させることなく音声認識率を向上させることが可能となる。

【0051】請求項 2 記載の発明によれば、入力音声の音声特徴量を抽出を比較的少ない計算量で得られ、かつ、音声認識率の向上に有効なパワー系特徴パラメータを抽出する頻度が、周波数軸系特徴パラメータを抽出する頻度よりも多く（自然数倍）なるので、処理速度を低下させることなく、音声認識率を向上させることが可能となる。

【0052】請求項 3 記載の発明によれば、同一の時間窓において抽出されたパワー系特徴パラメータと周波数軸系特徴パラメータとからなる従来の音声特徴量に対して、更に、それ以前の時間窓において抽出されたパワー系特徴パラメータを付加したものを新たな音声特徴量として用いるので、処理速度を低下させることなく、従来に比べてより精度良く音声認識を行うことが可能となる。

【図面の簡単な説明】

【図 1】本発明の一実施の形態における音声認識装置 1 の内部構成を示すブロック図である。

【図 2】時間窓位置設定部 11 における、入力音声波形に対する時間窓の設定の様子を示す図である。

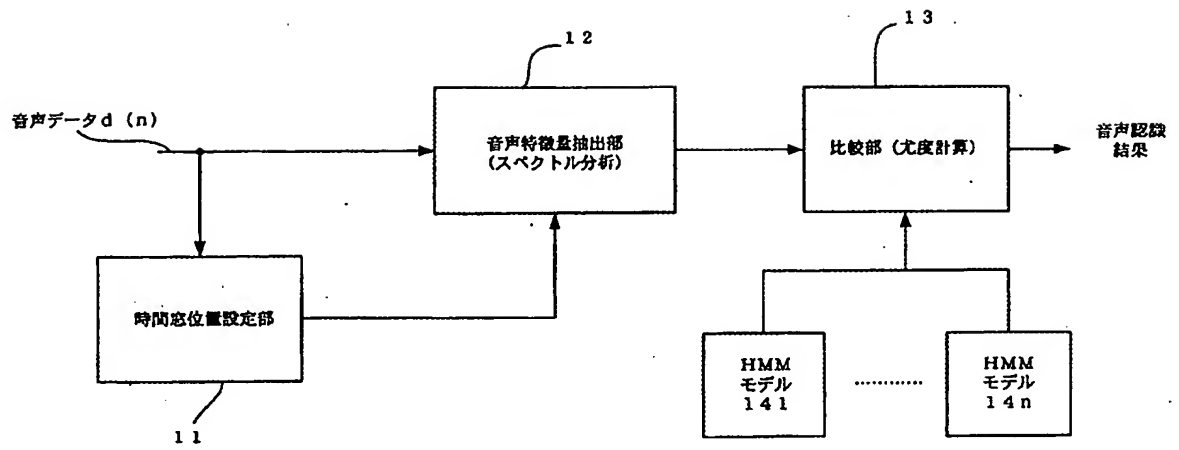
【図 3】本発明の一実施の形態における音声認識装置 1 の音声認識処理を説明するフローチャートである。

【図 4】従来技術における、入力音声波形に対する時間窓の設定の様子を模式的に示す図である。

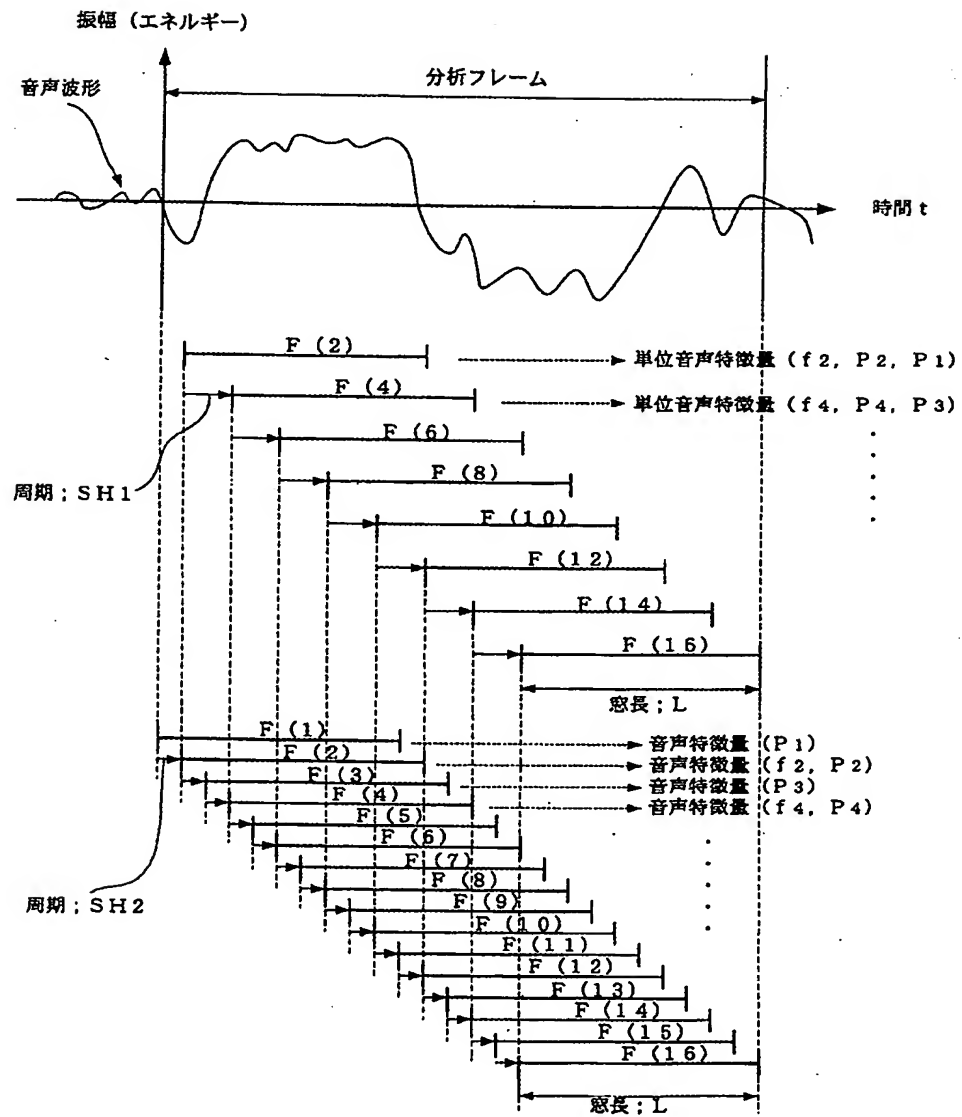
【符号の説明】

- 1 音声認識装置
- 11 時間窓位置設定部
- 12 音声特徴量抽出部
- 13 比較部
- 141～14n HMMモデル

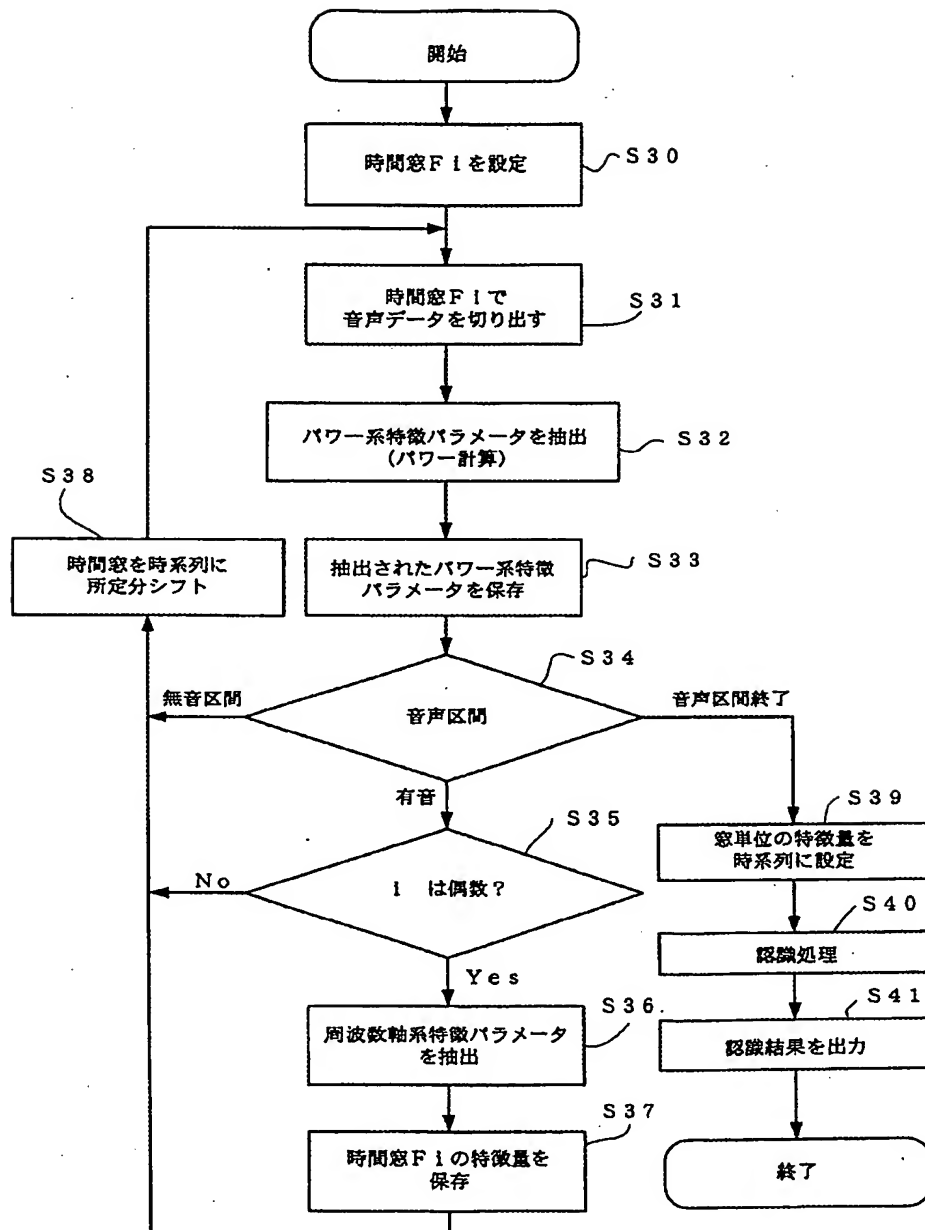
【図1】



【図 2】



【図 3】



【図 4】

